

## A Comparison of Two Analytic Methods for the Identification of Neighborhoods as Intervention and Control Sites for Community-Based Programs

By: Patricia O'Campo, Margaret O'Brien Caughy, Robert Aronson and Xiaonan Xue

O'Campo P, Caughy M, [Aronson R](#), Xue X. (1997). A comparison of two analytic methods for the identification of neighborhoods as intervention and control sites for community-based programs. *Evaluation and Program Planning*, 20(4): 405-414.

Made available courtesy of Elsevier: [http://dx.doi.org/10.1016/S0149-7189\(97\)00034-7](http://dx.doi.org/10.1016/S0149-7189(97)00034-7)

**\*\*\*Reprinted with permission. No further reproduction is authorized without written permission from Elsevier. This version of the document is not the version of record. Figures and/or pictures may be missing from this format of the document.\*\*\***

### **Abstract:**

*Study Objective:* Interest in community as the focus of public health interventions is growing. However, choosing intervention and comparison neighborhoods when designing community based programs poses a challenge to program planners. Ideally, intervention neighborhoods should be chosen based upon risk profiles and demonstrated need for the program. Multiple sources of data that tap into neighborhood characteristics might be used to facilitate the selection of intervention and comparison neighborhoods for program implementation and evaluation.

*Design:* We present and compare selected characteristics of two analytic methods that can be used to create perinatal risk profiles of neighborhoods within cities. For our example, we used information from several sources of routinely available data and used census tract level low birthweight as our intervention or outcome variable.

*Main Results:* At the neighborhood level, we found average household wealth of the census tract, proportion of births to women with late or no prenatal care, proportion of teen births per census tract, per capita crime rates, proportion of housing violations, and number of community organizations as being important factors identifying neighborhoods at risk for high rates of low birthweight births. Advantages of both methods are discussed and risk profiles generated from either method can be used not only to identify high risk areas of the city for adverse perinatal outcomes but also for the identification of intervention and comparison neighborhoods for implementation of community based programs.

**Keywords:** community interventions, evaluation design, control group, low birth weight, infant mortality, regression analysis

### **Article:**

#### **INTRODUCTION**

Interest in community as the focus of public health interventions is growing (Shea, 1992). Not only are the number of community-based public health interventions and programs on the rise (Worden et al., 1987; Pentz et al., 1989), there is also an increased recognition that characteristics of communities or neighborhoods<sup>1</sup> explain and can influence individual health behaviors and outcomes (O'Campo, Gielen, Faden, Xue, Kass, & Wang, 1995a; Brewster, 1994; Smith & Jarjoura, 1989; Garner & Raudenbush, 1991; Haan, Kaplan, & Camacho, 1987; Feldman, Makuc, Klineman, & CornoniHuntley, 1989; Marmot, Kogevinas, & Elston, 1987; Syme & Berkman, 1976; Cockerham, Kunz, & Lueschen, 1988; Tryoler & Cassel, 1964; Brooks-Gunn et al., 1993).

Often, when identifying determinants of adverse health outcomes the focus is on individual-level risk factors. In part, this is due to the abundance of public health studies that focus exclusively on determinants of health at the individual level (O'Campo et al., 1995a; Susser & Susser, 1996; Pierce, 1996). However, environments contribute directly and indirectly to adverse health. Community-level risk factors include environmental stressors which shape vulnerability and resistance to individual risk factors and protective factors for health (Cassel, 1976). For example, living in an impoverished, high crime area may elevate stress levels and increase susceptibility to ill health (Rutter, 1981; Rutter, Cox, Tupling, Berger, & Youle, 1975; Hindlin & Buchanan, 1993). Thus

neighborhood characteristics, in addition to characteristics of individuals, must be taken into consideration when characterizing sites for community based interventions.

Choosing intervention and comparison neighborhoods when designing community based programs poses a challenge to program planners and evaluators. Program planners need to be able to target areas where funds will have a substantial impact (e.g. Rutter et al., 1975; Hindlin & Buchanan, 1993; O'Campo, Guyer, Squires, Weiss, Sweitzer, & Coyle, 1993) and evaluators need to find communities of comparable risk as intervention neighborhoods. Ideally, intervention neighborhoods should be chosen based upon risk profiles and demonstrated need for the intervention program.

Previous neighborhood-level analyses have employed a wide variety of analytic techniques to describe at-risk communities. Often, researchers examine social risk factors such as high crime rates, poor housing conditions, separately by neighborhood (Rutter et al., 1975; Hindlin & Buchanan, 1993). A more advanced method of analyzing community-level risk, employed less frequently, is the simultaneous examination of neighborhood-level risk factors either by regression methods (Garbarino & Sherman, 1980; O'Campo, Xue, Wang, & Caughy, 1997; Iverson, 1991; Byrk & Raudenbush, 1992) or through creation of a composite risk score (O'Campo et al., 1993). Although use of multiple sources of data that tap into neighborhood characteristics strengthen such analyses, it is relatively infrequent that data other than census or vital records are used (O'Campo et al., 1993; Shevsky & Bell, 1955; Struening, 1983; Smith, 1979). Examples of routinely available data that could be used include census data; data from local governments on unemployment, housing, voting, and community group activity; and commercially available data on socioeconomic profiles of neighborhoods in cities.

We present and compare two analytic methods that can be used to create perinatal risk profiles of neighborhoods within cities. These methods, regression analysis and principal components analysis, have been used in past research to identify high risk neighborhoods (O'Campo et al., 1993; Garbarino & Sherman, 1980; Shevsky & Bell, 1955; Struening, 1983). The two types of social risk analyses, however, have not been compared previously in terms of whether they yield similar rankings of high risk neighborhoods. We use information from several sources of routinely available data. We demonstrate how the resulting risk profiles can be used to identify high risk areas of the city for implementation of interventions and for the selection of comparison neighborhoods for evaluation of such programs.

## **METHODS**

### ***Sources of Data and Variable Definitions***

Data at the census tract level were used for this research. Although zip code information is more readily available on routinely available data, zip codes are too large geographically and too heterogeneous with respect to risk to be useful for the purposes of identifying high risk neighborhoods. Thus, we used information from census tracts. Census tract boundaries are determined by the US Bureau of the Census and typically aggregate information on approximately 4000 persons. Baltimore City has 198 residential census tracts. Census tract data for the current research came from several sources that are described below.

### ***Baltimore City Vital Records***

We were interested in the latest data on low birth weight in Baltimore City. Thus, computerized birth certificates for the calendar years 1985-1989 were obtained from the local City Health Department Bureau of Biostatistics (in 1991 and early 1992, when these analyses were being conducted, data for 1990 were not yet available). These birth certificates contain information on birth weight and individual-level risk factors for LBW: maternal age, maternal education, timing of initiation of prenatal care, and whether the mother was eligible for medical assistance insurance. The local health department routinely geocodes data with census tract identifiers. For the 198 residential census tracts, births were aggregated to obtain rates and proportions of births for the following variables. Low birthweight was defined as infants born weighing < 2500 grams; teen birth was defined as births to women who were ≤ 17 years of age; and late prenatal care initiation was defined as prenatal care that was initiated later than the first trimester. The proportions of women giving birth in each census tract who had fewer than 10 years of schooling and who were eligible for medicaid were also obtained from vital records.

TABLE 1  
PROFILE OF BALTIMORE CITY CENSUS TRACTS ON SELECTED BIRTH, SOCIODEMOGRAPHIC AND NEIGHBORHOOD INFORMATION

	Mean (s.d.) or Median	Range
Low birth weight rate	0.11 (0.04)	0.006–0.20
Proportion teen births	0.20 (0.10)	0.007–0.50
Proportion of births to women with < 10 years of schooling	0.12 (0.09)	0.006–0.5
Proportion of births to women who initiated prenatal care $\geq$ 2nd trimester	0.10 (0.05)	0.007–0.5
Proportion of births to women eligible for medicaid	0.33 (0.18)	0.006–0.73
Ratio of home owners to renters	1.4 (1.30)	0.002–7.072
Ratio of abandoned to non-abandoned houses	0.02 (0.09)	0–1.315
Proportion of housing violations	0.09 (0.07)	0–0.41
% Families in poverty	22.68 (15.9)	1–79.21
Average household income (thousands)	24.0 (9.66)	4.9–64.98
Average household wealth (thousands)	95.4 (47.4)	8.90–287.83
Per capita crime	0.11 (0.18)	0.009–2.48
Median number of community groups	2	0–12

### *City Planning Data*

The Community Planning Division of the City of Baltimore's Department of Planning makes available data on a variety of neighborhood characteristics at the census tract level. In 1990, the Planning Division produced census tract profiles which included information on home ownership, numbers of abandoned houses, numbers of housing violations issued, percentage of families living in poverty, and per capita crime rates. The Planning Division also collects information on community groups that are registered and active in the City. Community groups ranged from crime watch and home owner associations to local political interest groups. This information is not geocoded, so attachment of census tract information was performed by the authors. Once geocodes were attached, the number of groups per census tract was tabulated.

### *Commercially Available Data*

There are sources of commercially available data that yield profiles of census tracts. Household wealth information (e.g., equity in homes, vehicles, assets, mortgages and other debts and real estate ownership) at the census tract and block group level can be obtained through Claritas/NPDC Inc. (P.O. Box 610, Ithaca, NY 14851-0610, U.S.A.). This information is updated annually based upon marketing surveys and projections. Household wealth and average household income were used as risk factor data for the current analytic models. Both wealth and income were categorized into 4 levels for the regression analyses. The four categories of wealth were \$0—\$50,000, \$51,000—\$100,000, \$101,000–\$125,000, and > \$125,000. Wealth is a better indicator of overall economic resources available to a family than annual income (Oliver & Sharpiro, 1995).

## **ANALYTIC METHODS**

### *Descriptive and Bivariate Analyses*

First, univariate distributions and correlations for each risk variable used in the analyses were generated. Means or medians and ranges and correlations for all 12 risk variables are presented in Tables 1 and 2. Bivariate analyses for the outcome low birthweight were then performed.

### *Binomial Logistic Regression*

Binomial logistic regression analyses were used to regress low birthweight births on the twelve independent census tract level variables. Specifically, for each of the 198 census tracts, the outcome was defined as number of low birthweight births divided by the number of total births for the census tract. Although this type of regression analysis is very similar to performing a linear regression on the proportion of low birthweight births per census tract, binomial logistic regression was our preferred method. Binomial logistic regression takes into account the exact variance for each census tract which varies by the number of total births for the census tract. For these regression analyses, model building techniques were employed. Specifically, covariates were added to the model

one by one, and at each step, a likelihood ratio test was performed to determine if the covariate contributed significantly to the model. Because the sample size was not large ( $N = 198$ ), the most parsimonious 'best-fit' model was desired. The predicted values of low birthweight from the best-fit regression model were used to rank the census tracts from highest to lowest low birthweight rates. The 40 census tracts with the highest ranks (or highest predicted levels of low birthweight) were identified as the two highest census tract quintiles of risk for low birthweight. The two highest quintiles are spatially displayed on a map in Figure 1. The ranks of the highest quintile of risk appear numbered 1-40 on the map where 1 is the census tract in the City that was at the highest risk low birthweight. Quintiles were appropriate for our purposes of the identification of two intervention and one comparison community of approximately eight census tracts each. Had we been choosing a smaller area for intervention and comparison communities we might have divided the city into sextiles or septiles.

TABLE 2  
CORRELATIONS BETWEEN RISK VARIABLES

	2	3	4	5	6	7	8	9	10	11	Community groups
1. % Teen births	0.57	0.65	0.86	-0.45	0.30	0.56	0.64	-0.64	-0.59	0.04	0.16
2. % < 10 Years of schooling		0.27	0.47	-0.24	0.16	0.31	0.43	-0.47	-0.43	0.06	0.01
3. % Prenatal care $\geq$ 2nd trimester			0.66	-0.45	0.30	0.48	0.52	-0.42	-0.46	0.12	0.12
4. Proportion births to women eligible for medicaid				-0.60	0.24	0.57	0.76	-0.62	-0.68	0.13	0.22
5. Ratio of home owners to renters					-0.18	-0.26	-0.62	0.40	0.72	-0.24	-0.16
6. Ratio of abandoned to non-abandoned houses						0.25	0.28	-0.15	-0.17	0.13	0.02
7. Proportion of housing violations							0.33	-0.35	-0.26	0.03	0.11
8. % Families in poverty								-0.56	-0.74	0.17	0.21
9. Average household income									0.80	-0.03	-0.15
10. Average household wealth										-0.25	-0.17
11. Per capita crime											0.09

### *Principal Components Analysis*

Principal components analysis was also performed to compare the results to that obtained from regression analysis. Principal components techniques provide a convenient means for creating a simplified summary measure from several items. Items that are part of a single factor are weighted and then combined to yield a score. In this case, the score reflects socioeconomic and reproductive disadvantage for each census tract. To facilitate direct comparison of the two methods we used the six census tract level risk variables from the best-fit regression model and low birthweight. The scores for each of the census tracts generated from the principal components analysis were ordered from worst to best. Census tracts falling into the two highest quintiles are displayed on the map in Figure 2.

### *Comparison of Two Analytic Methods*

We were interested in a specific comparison regarding the performance of regression analysis and principal components analysis. Specifically we were interested in how the two methods ranked the neighborhoods in terms of risk using the same information on risk factors'. We first identified important risk factors using the binomial regression analysis. The six factors identified as being important in predicting LBW were then also subjected to principal components analysis. Ranks of all census tracts were generated based upon the predicted values of LBW in the case binomial regression and the factor scores for the principal components analysis.

The ranks of the neighborhoods falling into the highest quintile of risk using regression were compared to the ranks for those census tracts obtained using factor analysis and nonparametric statistical tests were applied to the ranks generated by the two analytic methods.

## RESULTS

Table 1 displays the characteristics of the 198 residential census tracts in the City. The proportion of low birthweight per census tract ranged from a low of near zero to a high of 20% with an average low birthweight proportion for the City of 11 %.

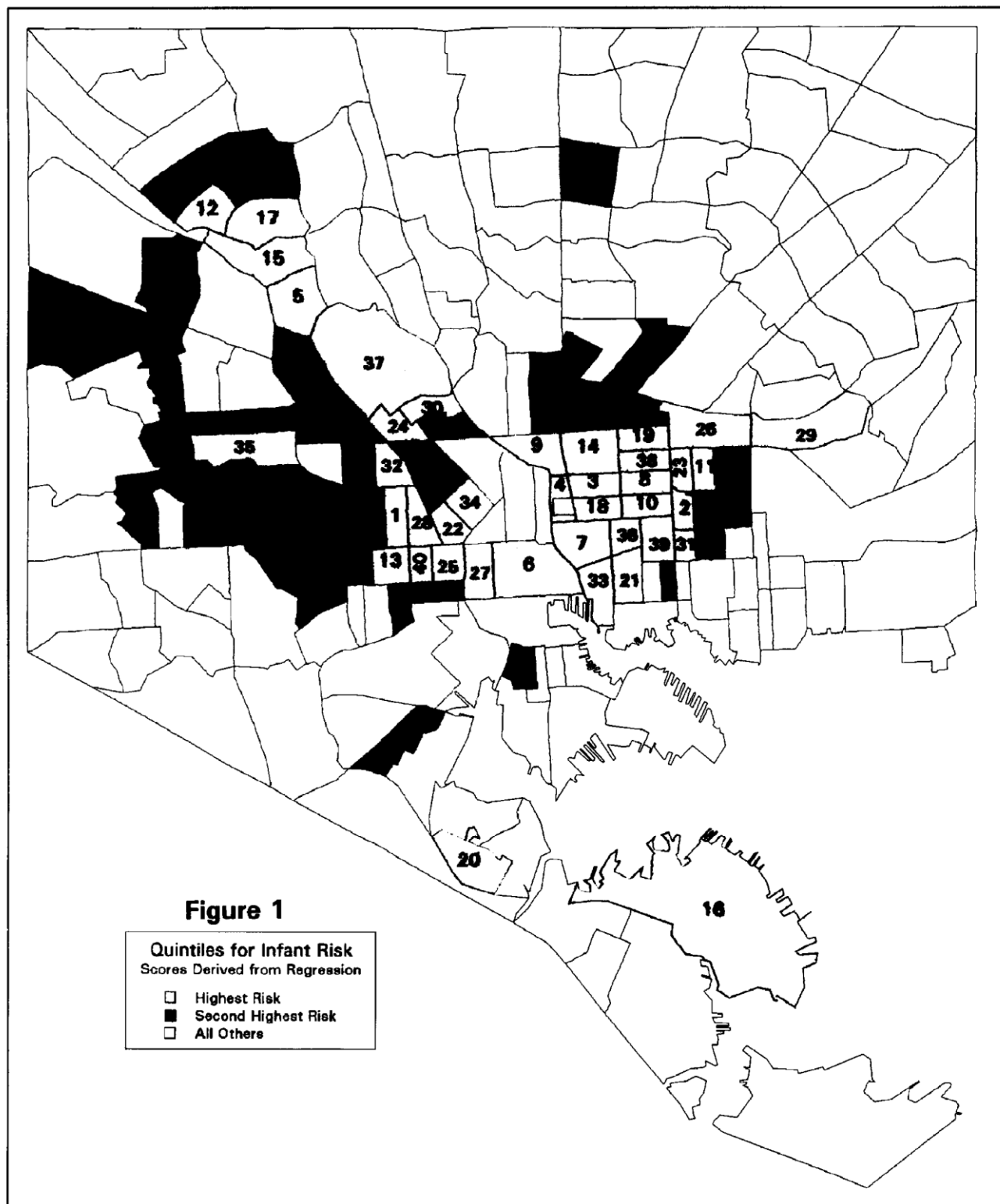


Figure 1. Quintiles of risk based upon results from regression analyses.

Pearson correlation coefficients for the twelve risk variables are displayed in Table 2. The proportion of teen births, late initiation of prenatal care, and eligible for medical assistance insurance are correlated. Poverty was highly correlated with average household income and average household wealth. Number of community groups per census tract was not highly correlated with any other risk factor.

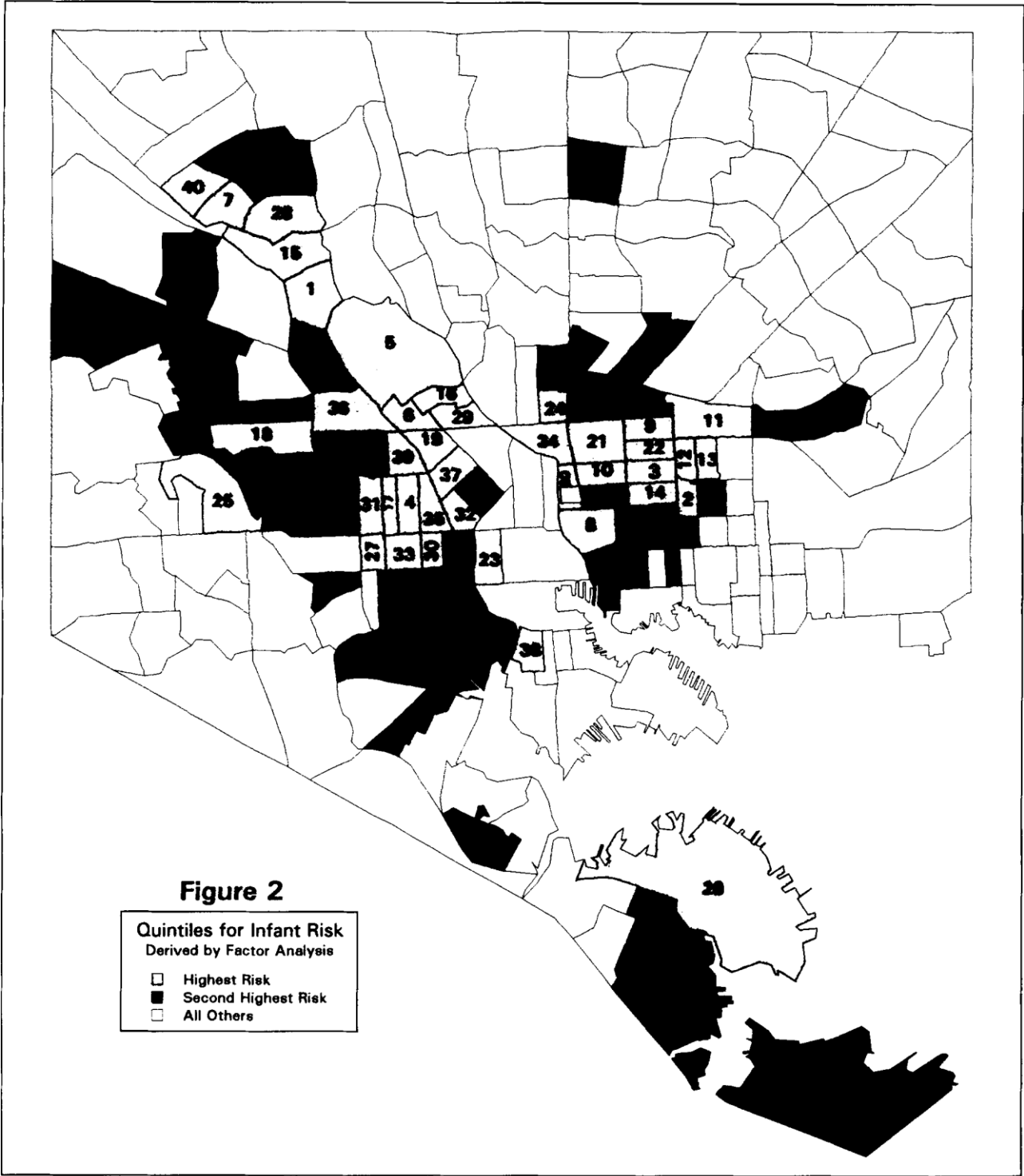


Figure 2. Quintiles of risk based upon results from factor analyses.

When bivariate analyses were performed, all risk factors showed associations with low birthweight that were statistically significant (Table 3).

Table 4 presents the best-fit regression model and includes the variables of average household wealth, initiation of prenatal care after the 2nd trimester, proportion of teen births, number of community groups, proportion of housing

violations and per capita crime. Because of the high correlations seen between poverty, income, and wealth, and the potential for multicollinearity, only one of these variables, average household wealth, was retained for the regression analysis.

TABLE 3.  
BIVARIATE BINOMIAL LOGISTIC REGRESSION ANALYSES FOR LOW BIRTHWEIGHT

	Beta (s.e. $\beta$ )	chi-square
Poverty	0.127 (0.0007)	0.0001
% Teen births	3.31 (0.1472)	0.0001
% Births to women with < 10 years school	1.05 (0.1394)	0.0001
% Births to women late prenatal care	6.22 (0.2985)	0.0001
% Births to women eligible Medicaid	1.85 (0.0738)	0.0001
Ratio home owners to renters	-0.19 (0.0117)	0.0001
Proportion abandoned houses	0.54 (0.0909)	0.0001
Proportion housing violations	2.19 (0.1500)	0.0001
Average household income	-0.00002 (0.000001)	0.0001
Average wealth	-0.0000048 (0.00000028)	0.0001
Community groups	0.032 (0.0038)	0.0001

TABLE 5  
COMPARISON OF RANKS: REGRESSION VERSUS PRINCIPAL COMPONENTS ANALYSES

		RANKS	
Regression	Prin. Comp	Regression	Prin. Comp
1	4	11	14
2	2	12	8
3	10	13	35
4	20	14	22
5	1	15	16
6	32	16	25
7	5	17	29
8	3	18	45
9	23	19	11
10	9	20	51

Regression median: 11.5.  
Princ. comp median: 15.

TABLE 4  
MULTIVARIATE BINOMIAL LOGISTIC REGRESSION ON LOW BIRTHWEIGHT: 'BEST-FIT' MODEL

Neighborhood Characteristic	Beta (s.e.)	chi-square
Wealth 50-99K <sup>1</sup>	-0.149 (0.03)	0.0001
Wealth 100-124K <sup>1</sup>	-0.191 (0.41)	0.0001
Wealth $\geq$ 125K <sup>1</sup>	-0.146 (0.5)	0.0023
Late Prenatal Care	2.816 (0.39)	0.0001
% Teen births	2.001 (0.21)	0.0001
Community groups	0.011 (0.004)	0.0045
% Housing violations	0.502 (0.180)	0.0054
Per capita crime	0.282 (0.118)	0.016

<sup>1</sup>The comparison group for wealth is <\$50,000.

Principal component analyses yielded a single factor with factor loadings for five of the items that ranged from a high of 0.89 for low birthweight and teen births to 0.67 for proportion of housing violations. Only number of community groups per census tract loaded low at 0.26.

Figures 1 and 2 depict the spatial distribution of the two highest quintiles identified from regression and principal component analyses, respectively. The actual ranks of the highest quintiles also appear on the maps.

There is much overlap in the results of the two methods, but the rankings of the neighborhoods, both in terms of quintile and actual ranking as indicated by the numbers, showed notable differences. We can see this if we specifically compare the rankings of census tracts falling into the highest quintile of risk as determined by regression and the corresponding rank as assessed through factor analysis (Table 5). A Kruskal-Wallis test for ranks (chi-square approximation) applied to the two ranking methods yielded a chi-square value of 387.86 ( $df$  197,  $p < 0.001$ ). This result indicates that the two ranking methods differ<sup>3</sup>. There were notable differences in the two methods, for example the fourth and the sixth highest risk neighborhoods according to regression were ranked 20 and 32, respectively (Table 5). Two neighborhoods in the highest quintile of risk using regression, 18 and 20, were ranked in the 2nd highest quartile of risk using factor analysis, 45 and 51 respectively (Table 5).

### *Use of Ranking to Identify Intervention and Comparison Neighborhoods*

Ranking of census tracts by risk can be used to identify neighborhoods in need of intervention programs as well as neighborhoods which may serve as comparison sites for evaluation. We will illustrate an example using the map

generated from the regression analysis (Figure 1). Intervention and comparison sites for relatively small programs (e.g., two census tracts) would be relatively easy to locate. Most likely, choosing these neighborhoods from within the highest quintile of risk within the City would be desirable. For a very high risk pair of adjacent census tracts numbers 3-4 and 6-7 could be used for intervention and comparison neighborhoods. For moderately high risk pairs of sites, the adjacent census tracts of 40-25 and 36-21 could be used. If a larger intervention site was needed (e.g., eight census tracts) then the two adjacent census tract clusters of 94-14-3-19-38-8-23 & 18-7-10-36-21-39-2-31 could be used for the intervention of comparison sites. If the intervention and comparison sites should not be located next to each other then the cluster of 1-13-26-40-2225-34-27 could be used along with one of the clusters mentioned above with the cluster 18-7-10-36-21-39-231 probably being more comparable with respect to risk.

## **DISCUSSION**

We sought to compare two analytic methods in terms of identifying neighborhoods in need of intervention programs based upon characteristics of sociodemographic and reproductive risk. Although the two methods, regression analyses and principal components analyses, yielded similar results in the identification of quintiles of risk, there were notable differences in the rankings of individual census tracts. One possible explanation is that not all variables that were important in the regression analysis yielded high factor loadings in the principal components analysis. The number of community groups per census tract, for example, had a low factor loading in the principal components analysis. Factor analysis performs best when all items load high (i.e., all items are highly correlated with one another). That may, in fact, be a disadvantage of principal components analysis in that it may be useful for the purposes of needs assessment or program planning to use many characteristics of neighborhoods in choosing intervention and comparison sites. To require that all variables of interest to be highly intercorrelated may defeat the purpose of using a wide variety of factors.

On the other hand, an advantage of principal components analysis is that it can be easily used when there are multiple outcomes of interest (given that they are correlated). There are instances when program planners may not want to target a program to one outcome such as LBW in our example. Rather there could be a more general program of LBW and teen pregnancy prevention, or teen pregnancy and high school dropout prevention. In those cases, using principal components analysis would be advantageous in that all outcomes of interest would be included in the factor identification and creation process. Regression analysis explicitly identifies an outcome variable, low birthweight in this case. For the results of principal components to be of use, the outcome variable must load high in the analysis. In our example low birth weight did have a high loading. Because we were interested in a single outcome and preferred an analytic method that explicitly identified the outcome, we felt that the regression analysis was more appropriate for our purposes of identifying neighborhoods as intervention and comparison sites for a program targeting the prevention of LBW.

There are other practical differences between the two methods that might be noted. First, regression analysis is a statistical model that carries with it a number of model assumptions and requirements regarding the determination of goodness of fit. Thus, regression analysis can only be valid when the model assumptions can be met and the fit of the model is good. Although, similar requirements regarding model assumptions are not present with the use of principal components analysis, this method requires all variables that are included in a factor to be highly correlated as mentioned earlier. In general, it is important to keep in mind that when selecting an analytic method for identification of high risk areas, the advantages and disadvantages of the methods as well as how these advantages and disadvantages apply to the problem under study should be considered.

Some limitations of our analytic methods should be noted. The utility of ecological analyses, as presented here, are primarily for identification of high risk areas. Interpretation of the parameter estimates from macro-level logistic regression should be made with caution. Analyses that are performed at the group level almost always yield stronger relationships than would be observed if the same analyses were conducted with data at the individual level because the outcome variable reflects variation at the individual-level and the neighborhood-level data does not account for individual differences within the neighborhoods that may contribute to the variation in the outcome (e.g., health behaviors such as smoking) (Byrk & Raudenbush, 1992; O'Campo, Gielen, & Davis, 1995b). Thus, the parameter estimates, or betas, cannot be directly interpreted as the risk estimates for individual-level low



birthweight. We can, however, determine from both types of analyses the neighborhood level factors that help to explain variation in census tract level low birthweight rates.

Our analyses focus only on census tract as the unit of analysis. Previous analyses have used various area level units such as zip codes, SMSAs and census block groups. Although census tracts are probably preferable to zip codes or SMSAs because they are the smaller unit of analysis, census block groups might be the most homogenous unit of analysis because of its small size of approximately 10 city blocks. We chose to use census tract since program data are often geocoded for census tract and to our knowledge are not coded with census block group numbers.

Our findings suggest that use of several sources of data (i.e., vital record, census, city planning and commercially available data) to identify high risk areas of the city for adverse perinatal outcomes is useful. Three variables—average household wealth, community groups and housing violations—not often used in needs assessments, were significant in identifying high risk neighborhoods in the regression analysis. Use of supplemental routinely available data for needs assessments and risk profile analyses are encouraged as these sources are widely accessible to program planners.

Average household wealth was the most important predictor in the regression analyses of low birthweight risk. Use of information on wealth may be preferable to more traditional measures such as poverty, educational levels and income for several reasons. Poverty is limited in that it only identifies groups that have inadequate resources for daily living. Although this is an important predictor for low birthweight, other potentially high risk groups who are not living in poverty (e.g., those living at 130% of poverty) are not captured by this information. Moreover, because of the eligibility criteria for some public programs, those just above poverty may be worse off for some resources such as prenatal health insurance than those living in poverty. Income reflects current availability of economic resources, whereas wealth captures additional information on savings and assets which may reflect cumulative information on socioeconomic standing. Furthermore, when looking at risk by race or ethnicity, different race and ethnic groups have been shown to have wider variation in wealth than in education or annual income (Oliver & Sharpiro, 1995). Specifically, recent national survey data show that among families considered middle class, differences in median wealth of black and white families may be far greater than median income differences (Oliver & Sharpiro, 1995). Specifically, the Black/White (B/W) ratio of median family income was 1.6 (\$25,000 to \$15,630) while the B/W ratio of median wealth of those same families was 11.8 (\$43,800 to \$3700) (Oliver & Sharpiro, 1995). Given that researchers, program planners and evaluators need to do a better job of accounting for social differences between racial and ethnic groups, wealth may be a better choice of a socioeconomic indicator than the more traditional measures (Muntaner, Nieto, & O'Campo, 1996).

Information on housing violations and crime were also important predictors of low birthweight. Neighborhoods with high crime rates may cause residents to reduce outside social contacts which can lead to isolation. Even among residents with supportive networks, fear of crime in these neighborhoods may lead to increased stress that can counteract any buffer that social supports provide.

Poor housing conditions, from structural deficiencies to location of housing, have been linked to a variety of adverse health outcomes (Smith, 1990). Smith, in a review of effects of housing and health, reported a number of studies that have found a range of poor health outcomes from respiratory symptoms and headaches to infections and poor mental health, especially among women (Smith, 1990). Our findings of housing violations as being an important explanatory variable in census tract level low birthweight rates is consistent with previous findings. The housing problems that are responsible for the violations (e.g., space, sanitation, structural deficiencies) may result in stressful and possibly unstable living environments. Alternatively, housing violations, in this case, may be serving as a proxy for poverty since this variable was correlated with teen births and medical assistance eligibility. However, even when wealth was in the model, housing violations were significant and, thus, is an independent contributor to risk of low birthweight.

Another factor that was important in predicting low birthweight that is not typically considered in studies of adverse perinatal outcomes is the presence of community associations and organizations. Since the number of

community groups per census tract was not strongly associated with any of the other risk factors, the positive association between community groups and low birth weight is likely to be indicative of other neighborhood features/characteristics not captured by the other variables. Community groups have been described as having a periodicity of presence in a community, in response to the rising and falling of concern about specific issues (O'Brien, 1993). When neighborhoods deteriorate the numbers of community groups may rise in an organized effort to bring about change. For instance, block watch and neighborhood watch groups emerge in response to perceived threats to safety that may be ignited by specific events. The number of community groups in a census tract may be an indicator that the perceived quality of life for community members is unacceptably low.

Future analyses aimed at identifying neighborhoods at risk for adverse outcomes should tap a variety of data sources. Both methods that we compared are easy to implement and would be useful for identifying intervention sites and comparable control communities but regression analysis may be the preferred method.

## References:

- Brewster, K. L. (1994). Race differences in sexual activity among adolescent women: the role of neighbourhood characteristics. *American Sociological Review*, 59, 408-424.
- Brooks-Gunn, J., Duncan, G. J. & Klebanov, P. K. (1993). Do neighborhoods influence child and adolescent development? *AJS*, 99, 353-395.
- Byrk, A. A., & Raudenbush, S. W. (1992). *Heirarchical linear models: Application and data analysis methods*. Newbury Park, CA: Sage Publications.
- Cassel, J. (1976). The contribution of the social environment to host resistance. *American Journal of Epidemiology*, 104, 107-123.
- Cockerham, W., Kunz, G., & Lueschen, G. (1988). Social stratification and health lifestyles in two systems of health care delivery: a comparison of the United States and West Germany. *Journal of Health and Social Behavior*, 29, 113-126.
- Feldman, J., Makuc, D., Klineman, & Cornoni-Huntley, J. (1989). National trends in educational differentials in mortality. *American Journal of Epidemiology*, 129, 919-933.
- Garbarino, J., & Sherman, D. (1980). High-risk neighborhoods and high-risk families: The human ecology of child maltreatment. *Child Development*, 51, 188-198.
- Garner, C. L., & Raudenbush, S. W. (1991). Neighborhood effects on educational attainment: a multilevel analysis. *Sociology of Education*, 64, 251-262.
- Haan, M., Kaplan, G., & Camacho, T. (1987). Poverty and Health: prospective evidence from the Alameda County Study. *American Journal of Epidemiology* 125, 989-998.
- Hindlin, R., & Buchanan, D. (1993). Data for Empowerment! The use of small area analysis in guiding and targeting program development. *Presentation at the 121st Annual Meeting of the American Public Health Association*, San Francisco, CA.
- Iverson, G. (1991). *Contextual analysis*. Newbury Park, CA: Sage Publications.
- Marmot, M., Kogevinas, M., & Elston, M. (1987). Social/economic status and disease. *Annual Review of Public Health*, 8, 111-135.
- Muntaner, C., Nieto, F. J., & O'Campo, P. (1996). The Bell curve: On race, social class and epidemiologic research. *American Journal of Epidemiology*, 144.
- O'Brien, W. (1993). Network effect. *The Neighborhood Works*, 16, 33-34.
- O'Campo, P., Guyer, B., Squires, B., Weiss, J., Sweitzer, J., & Coyle, T. (1993). Needs assessment for reducing infant mortality: The Baltimore City Healthy Start Program. *Southern Medical Journal*, 86, 1342-1349.
- O'Campo, P., Gielen, A., & Davis, M. (1995). Smoking cessation interventions for pregnant women: Review and future directions. *Seminars in Perinatology*, 19, 279-285.
- O'Campo, P., Gielen, A., Faden, R., Xue, N., Kass, N., & Wang, M. C. (1995). Violence by male partners against women during the childbearing year: A contextual analyses. *American Journal of Public Health*, 85, 1092-1097.
- O'Campo, P., Xue, X., Wang, M. C., & Caughy, M. (1997). Multilevel analysis of neighborhood-level risk factors for low birthweight in Baltimore City. *American Journal of Public Health*, 87, 1113-1118.

- Oliver, M., & Sharpiro, T. (1995). Race, wealth and inequality in America. *Poverty and Race Research Action*, 4, 1-4.
- Pentz, M., Dwyer, J., Mackinnon, D., Flay, B., & Hansen, (1989). A multi-community trial for primary prevention of adolescent drug abuse. Effect on drug use prevalence. *Journal of American Medical Association*, 261, 3259-3266.
- Pierce, N. (1996). Traditional epidemiology, modern epidemiology, and public health. *American Journal of Public Health*, 86, 678-683.
- Rutter, M. (1981). The city and the child. *American Journal of Orthopsychiatry*, 54, 610-625.
- Rutter, M., Cox, A., Tupling, C., Berger, M., & Youle, W. (1975). Attainment and adjustment in two geographical areas. I. The prevalence of psychiatric disorder 1975. *Review of Child Development Research* (Vol. 4). Chicago: University of Chicago Press.
- Shea, S. (1992). Community Health, Community Risks, Community Action. *American Journal of Public Health*, 82, 785-787.
- Shevsky, F., & Bell, W. (1955). *Social area analysis: Theory, illustrative application and computational procedures*. Stanford, CA: Stanford University Press.
- Smith, D. A., & Jarjoura, G. R. (1989). Household Characteristics, neighborhood composition and victimization risk. *Social Forces*, 68, 621-640.
- Smith, N. (1979). Techniques for the analysis of geographic data in evaluation. *Evaluation and Program Planning*, 2, 119-126.
- Smith, S. (1990). Health status and the housing system. *Social Science and Medicine*, 31, 753-762.
- Struening, E. L. (1983). Social area analysis as a method of evaluation. In E. L. Struening & M. Brewer (Eds.), *Handbook of evaluation research*. Beverly Hills, CA: Sage Publications.
- Susser, M., & Susser, E. (1996). Choosing a future for epidemiology. II. From box to chinese boxes and eco-epidemiology. *American Journal of Public Health*, 86, 674-677.
- Syme, S., & Berkman, L. (1976). Social class, susceptibility and sickness. *American Journal of Epidemiology*, 104, 1-8.
- Tryoler, H., & Cassel, J. (1964). Health consequences of culture change II: the effect of urbanization on coronary heart mortality in rural residents. *Journal of Chronic Diseases*, 17, 167-177.
- Worden, J., Flynn, B., Solomon, L., Costanza, M., Foster, R. et al. (1987). A community-wide breast self-exam education program. In *Advances in cancer control: the war on cancer-15 5 years of progress* (pp. 27-37). New York: Liss.

## Notes:

<sup>1</sup>Neighborhood, in this paper will be used to denote areas defined by geographical boundaries, whereas community will be used to identify areas of social interaction (e.g. African—American business community, a workplace) which may coincide with but is not necessarily limited to a geographic neighborhood area.

<sup>2</sup>For our purposes, we were not interested in assessing whether the two methods might identify the same risk factors as being most important, although this is another component of the comparison that could have been examined. To accomplish this we would have had to put all 12 risk factors into the both the regression and principal components analysis to obtain two separate 'best fit' models. The results of the regression and the principal components analysis in terms of the variables identified as being significant would then be compared.

<sup>3</sup>The Kurskal—Wallis test is designed to be applied to two independent samples. In our case, the two methods of ranking were applied to the same data set. Therefore, the ranks are not independent. The problem presented by this lack of independence in the calculation of the significance test is not likely to be a large problem in our case. Specifically, it is not likely that the dependence of the two samples will bias the observed p value of <0.001 to the extent that the 'true' p value is > 0.05.